A Ground-Truth Video Dataset for the Development and Evaluation of Vision-based Sense-and-Avoid systems

Adrian Carrio, Changhong Fu, Jesus Pestana, and Pascual Campoy

I. INTRODUCTION

The use of fixed-wing Unmanned Aerial Systems (UAS) for civil tasks is recently becoming more and more important. Fixed-wing UAS can be useful for surveillance tasks, road traffic control, fire-fighting, meteorological observations, telecommunications, etc. Currently a lot of efforts are being put in developing UAS that are able to coexist safely and effectively with current manned operations in the national and international airspace [1]. UAS are expected to perform Sense and Avoid (SAA) functions at an "equivalent level of safety" (ELOS) to manned aircraft while not negatively impacting the existing infraestructure and manned Traffic Alert and Collision Avoidance System (TCAS) that create today's safe airspace [2] [3].

Active sensors like radar are not feasible due to power consumption and size restrictions onboard UASs, while vision provides a passive, low cost and low power solution.

The availability of adequate image datasets is essential for developing and evaluating vision-based aircraft detection and tracking algorithms. Nevertheless, this type of images are scarce and in general expensive to obtain. Furthermore, when dealing with the collision course detection problem, obtaining this type of images becomes a risky challenge.

Mian et al. [4] used images recorded from the ground for aircraft tracking purposes. Mejias et al. [5], [6], [7] and Dey et al. [8] describe different vision-based collision avoidance systems for which real flights were performed in order to collect suitable test data. To the author's knowledge, none of these test data have been made available for further development by the research community.

Currently some specific internet platforms can be found from which real aircraft imagery and videos can be downloaded such as AirTeamImages [9] or JetVideos [10]. However the imagery provided in these websites is not intended for algorithm development or testing and therefore many useful features are not available, e.g. georeferenced or annotated image(s) or the possibility of selecting images by different criteria: type of background, aircraft trajectory, etc.

Recent research on the use of simulated images for analyzing and demonstrating the functionality of vision-based systems has been performed by Zsedrovits et al [11], [12], where the flight simulation software FlightGear [13] has been used to obtain simulated flight imagery for which the intruder's size and 3D positions are known.

In order to improve the reality of this fully simulation-based approach, we have developed specific software to generate simulated images in which the virtual background is replaced with real-world backgrounds recorded from UAVs, taking into account visualization parameters, such as illumination, and on board camera vibrations. To the best of our knowledge, the development and use of this type of imagery for designing and evaluating vision-based Sense-and-Avoid systems is completely novel.

This paper is organized as follows: Section II introduces the developed software for video generation with simulated intruders. This software performs image processing to combine imagery captured from real UAV flights with 3D virtual imagery. In Section III a video dataset with ground truth data generated with our software is presented. This video dataset is public and the videos are freely available for download in Internet, allowing not only to develop, test and validate many different aircraft detection and tracking algorithms, but also to effectively measure their performance and have them compared within a common testbed. Our imagery is applied for evaluating different tracking algorithms in Section IV. Finally, section V presents the conclusions.

II. SIMULATED VIDEO GENERATION

Specific software has been developed allowing a fast and flexible generation of videos for Sense-and-Avoid systems'

development and evaluation. This software allows the simulation of different intruders with various possible aircraft models and user-set simulated trajectories. Also scene illumination parameters can be controlled by the user.

In order to create these simulated images, both real and virtual images are obtained and then combined through an image processing algorithm, taking into account on board camera vibrations.

A. DATA COLLECTION

Video imagery of real flights recorded from an USol K50 UAV was provided by Unmanned Solutions (USol[1]). These videos have been recorded with a GoPro Hero 2 HD camera located in the UAV tail. The camera uses 170° fish-eye lenses, capturing frames with a 1280×960 pixels resolution.

These videos offer a variety of illumination and cloud conditions which are very challenging for detecting or tracking aircrafts, as shown in Figure 1.

B. VIRTUAL IMAGES WITH SIMULATED INTRUDER

Virtual images were generated using a rendering library based on OpenGL. This library allows to render a 3D scene with realistic models of different aircrafts. The aircraft models used are GPL licensed and were downloaded from FlightGear's website [13], where different 3D models for gliders, propeller aircrafts and jets are available.

For the development of our dataset we have used Cessna 172 and Boeing 737 3D models shown in Figure 2 since these are some of the most common aircrafts in the world.

The intruder aircraft flight trajectory is defined out of a set of 3D points. A spline interpolation between the 3D points is computed and sampled for every frame in the background video. This 3D coordinates are used to define the position and orientation of the intruder aircraft in the 3D scene for each frame.

The 3D scene is rendered with a chroma key green backdrop so that the pixels corresponding to the intruder aircraft can be easily segmentated during the image processing stage.

The 3D scenes have been rendered using a NVIDIA GeForce GTX 675MX Graphics Card, with optimized rendering options including:

- Averaged normals for a smooth appearance
- Antialiasing to minimize the "ladder" effect
- Anisotropic filtering to improve texture sharpness

The camera for the rendered 3D scene is defined at a constant altitude of 150m and with a similar pose with respect to the UAV as that the real camera originally had. The homography transformations performed later on will modify the resulting rendered image according to the real camera movements.

A directional light source has been used to render the intruder. This light source is infinitely distant, such that its rays are all parallel to a direction vector. This vector has been estimated from the images and is specified for every background video, so that the illumination conditions on the intruder are as realistic as possible.

[1]http://www.usol.es

Fig. 1: Real flight video captures used as background. The videos were recorded during 2011 in Marugán (Spain) from an USol K50. The images show different cloud conditions.

C. IMAGE PROCESSING ALGORITHM

The image processing algorithm, developed with OpenCV [14], perfoms the compositing of the virtual image containing the simulated intruder and the background real image.

As previously noted the camera onboard the K50 is always located in the UAV's tail and suffers from intense mechanical vibrations. Therefore the objective of the compositing process is not only to place the intruder in the background image but to simulate the camera vibration effect and to have the intruder aircraft edges interpolated with the background in order to obtain a realistic appearance of the intruder.

First, we search for strong corners in the first frame of the

Fig. 2: Boeing 737 and Cessna 172 models used to simulate intruder aircrafts.

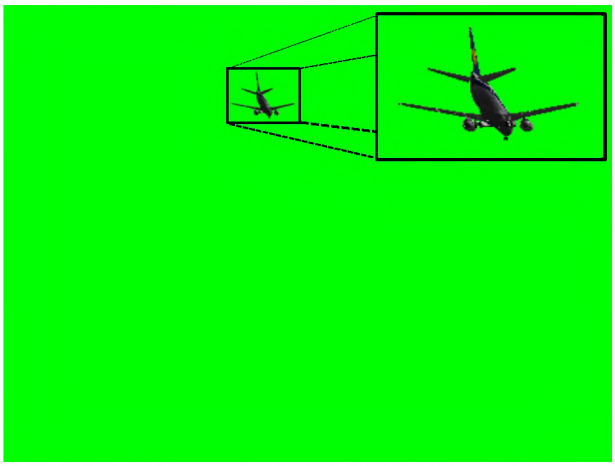

Fig. 3: Virtual image generated with a 3D rendered model over a chroma key green backdrop.

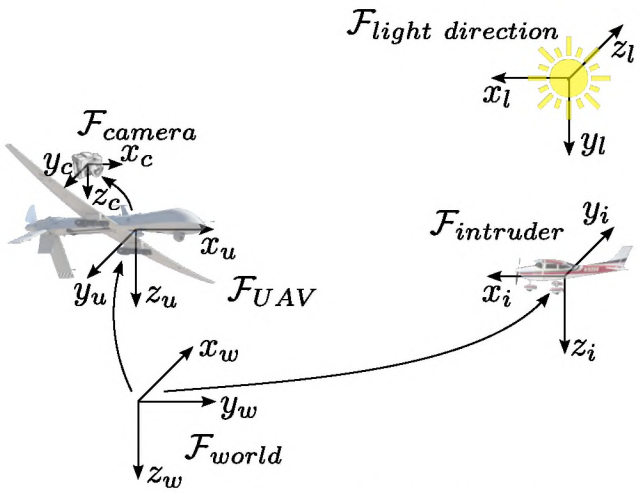

Fig. 4: Reference systems used for describing the camera, UAV and intruder poses and the light direction in the 3D scene.

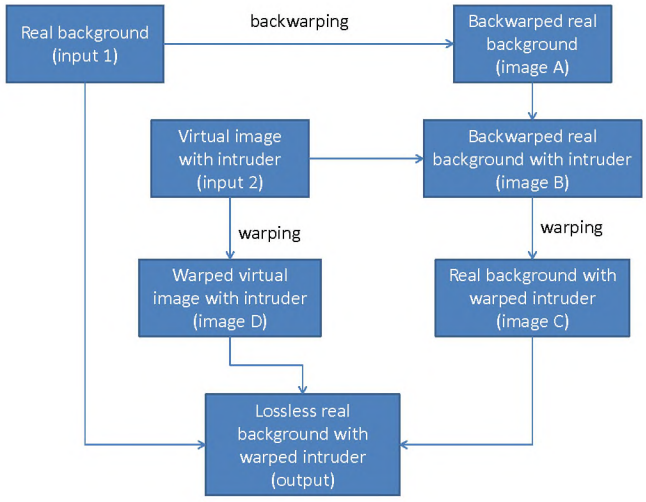

Fig. 5: Image compositing process. The output image is obtained compositing "input 1" and "image C" by using "image D" as a mask.

background video sequence using the GoodFeaturesToTrack method [15]. Since the UAV carrying the camera is itself visible in the images, specific masks have been obtained in order to skip the search for strong corners in this particular area of the image. The reason for doing this is that the perspective transformation will be estimated better with points belonging to a far plane.

The strong corners found are then matched in the following frames using the iterative Lucas-Kanade method with pyramids [16]. Once the matched points in the first frame (x_i, y_i) and the current frame (x_i', y_i') are obtained, a per-

Fig. 6: Masks developed for robust homography estimation. The areas colored in black are not considered for the search of strong corners.

spective transformation H between both frames is computed using a RANSAC-based robust method [17] so that the backprojection error (Eq. 2) is minimized. The perspective transformation H shown in Eq. 1 links coordinates between two views of the same scene.

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \quad (1)$$

$$\sum_i \left(x'_i - \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}} \right)^2 + \left(y'_i - \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}} \right)^2 \quad (2)$$

Once the perspective tranformation H is known the image compositing process shown in Figure 5 is followed. To obtain image A the real background is backwarped, that is, the perspective transformation H^{-1} given by Eq. 3. The values of pixels with non-integer coordinates are computed using a Lanczos interpolation over an 8x8 pixel neighborhood.

$$dst\,(x, y) = src\left(\frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}, \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \right) \quad (3)$$

Then the rendered pixels corresponding to the intruder are copied into the backwarped background (image B). In the next step, the obtained image B is warped, that is, the perspective transformation H is applied to obtain image C. In this frame the intruder will appear warped and its edges will be interpolated with the background. However, the perspective transformation might have caused a loss of information belonging to the real backgroung due to the interpolation process.

To fix this issue, the perspective transformation H is applied to the virtual image and image D is obtained. This image is used as a mask that indicates which pixels correspond to the background (green) and which to the warped intruder and its edges (not green).

Fig. 7: Close view of a simulated intruder. Edges appear interpolated with the background increasing the realistic appeareance of the intruder.

In order to build the output frame, the pixels belonging to the warped intruder and its edges according to image D, will be copied from image C into the real background.

The final result shows the unaltered real background and the intruder having suffered a perspective transformation, that is, a change of perspective due to onboard mechanical vibrations. The intruder's edges appear interpolated with the background, creating a very realistic effect, as shown in Figure 7.

D. COMPARISON WITH SIMILAR REAL-WORLD IMAGES

A simulated intruder has been placed next to a real intruder in various video sequences using the developed software in order to compare their appareance.

In these sequences, a light aircraft appears in the field of view of the UAV flying different trajectories, which we have tried to imitate while keeping both real and simulated aircrafts visible.

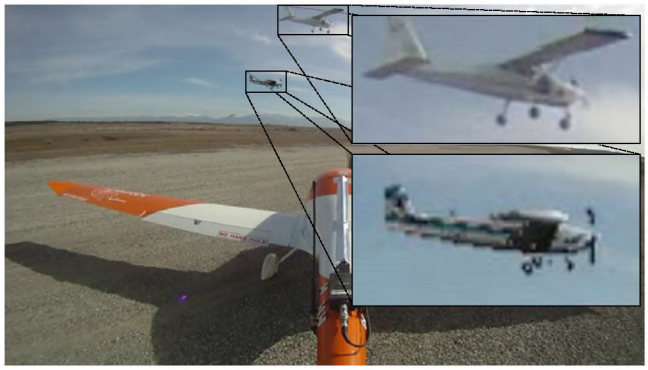

Fig. 8: Real (above) and simulated (below) intruders. Distance between camera and intruder is 58 m.

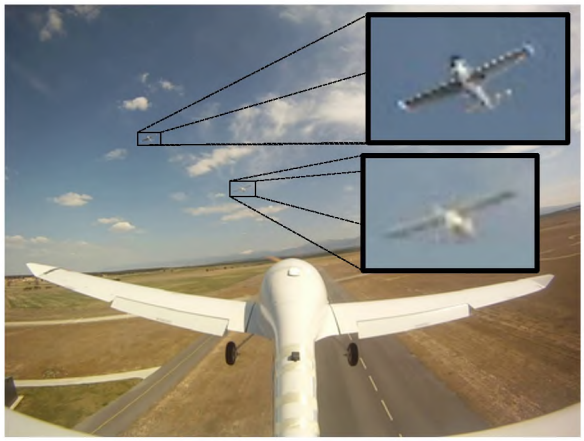

Fig. 9: Simulated (above) and real (below) intruders. Distance between camera and intruder is 156 m.

E. VIDEO ANNOTATIONS

Complete video annotations are included in XML format together with the video sequences including for every frame:

- Intruder 3D coordinates wrt to the camera
- Euclidean distance between camera and intruder
- Intruder bounding box in the image (top-left pixel coordinates, width and height)

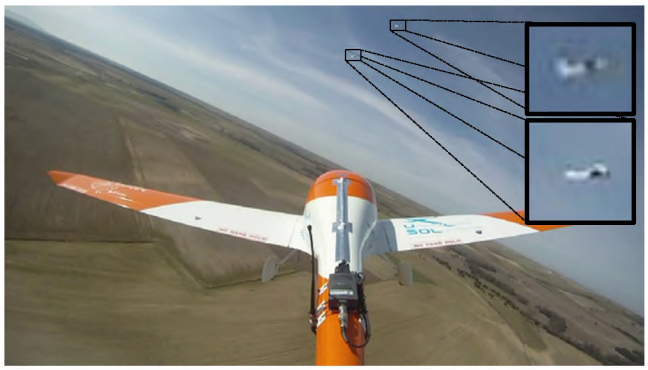

Fig. 10: Real (above) and simulated (below) intruders. Distance between camera and intruder is 333 m.

TABLE I: Video dataset

Filename	Intruder aircraft	No. frames	Trajectory
1LRB737.avi	Boeing 737	322	Left to right
1FNB737.avi	Boeing 737	389	Approaching
1NFB737.avi	Boeing 737	636	Moving away
1RLB737.avi	Boeing 737	275	Right to left
2LRB737.avi	Boeing 737	337	Left to right
2FNB737.avi	Boeing 737	318	Approaching
2NFB737.avi	Boeing 737	1143	Moving away
2RLB737.avi	Boeing 737	363	Right to left
3LRB737.avi	Boeing 737	353	Left to right
3FNB737.avi	Boeing 737	1033	Approaching
3NFB737.avi	Boeing 737	1033	Moving away
3RLB737.avi	Boeing 737	1033	Right to left
1LRC172.avi	Cessna C172	321	Left to right
1FNC172.avi	Cessna C172	411	Approaching
1NFC172.avi	Cessna C172	636	Moving away
1RLC172.avi	Cessna C172	274	Right to left
2LRC172.avi	Cessna C172	342	Left to right
2FNC172.avi	Cessna C172	287	Approaching
2NFC172.avi	Cessna C172	1143	Moving away
2RLC172.avi	Cessna C172	347	Right to left
3LRC172.avi	Cessna C172	337	Left to right
3FNC172.avi	Cessna C172	1033	Approaching
3NFC172.avi	Cessna C172	1033	Moving away
3RLC172.avi	Cessna C172	1033	Right to left

The reference systems defining the camera and intruder 3D positions are described in Figure 4. Units are defined in meters (m).

III. VIDEO DATASET

The video dataset currently consists of a set of 24 videos listed in table I. As previously mentioned, a set of real UAV flight videos was selected in order to provide a variety of trajectories, illumination and cloud conditions. The video dataset is freely avaliable for download at Vision4UAV [18] together with the associated ground-truth data in XML format. The sequences are provided free of charge for academic research. For any other use, please contact the authors. Should you care to publish these sequences or results obtained using, please indicate their origin as "CVG-UPM / EVision II Project", and mention the video dataSet webpage address [18].

IV. RESULTS

In this section, we have applied AM^3 [19] state-of-art visual tracker to video sequence 1RLB737.avi in order to check and evaluate the effectiveness of our dataset. The ground truth has been used to measure the tracking performance in terms of the X and Y comparison with the ground truth data.

The background of this video contains four main challenging factors: (1) Strong motions or large displacements, (2) illumination variation, (3) background clutters and (4) scale changes. Tracking results are shown in Figure 11.

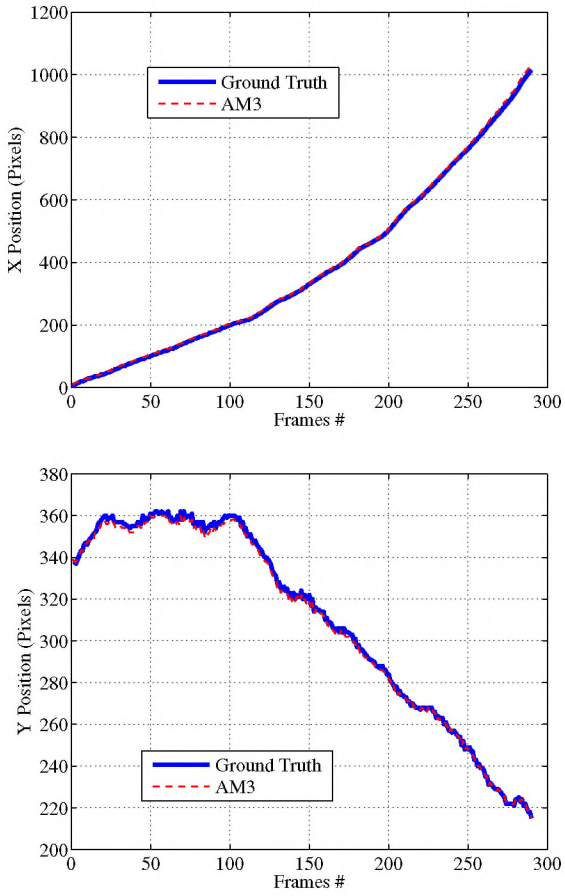

Fig. 11: X and Y Position comparison with Ground Truth data in 1RLB737.avi

V. CONCLUSIONS

Specific software has been developed for obtaining realistic simulated images of intruder aircrafts in real flight imagery. These images are created by compositing a 3D rendered intruder aircraft image with real flight images recorded from a UAV, taking into account illumination and on board camera vibrations. The images included in our video dataset can be used as a powerful benchmark for the development and evaluation of vision-based Sense-and-Avoid systems, therefore representing an important improvement in the state of the art for Sense-and-Avoid.

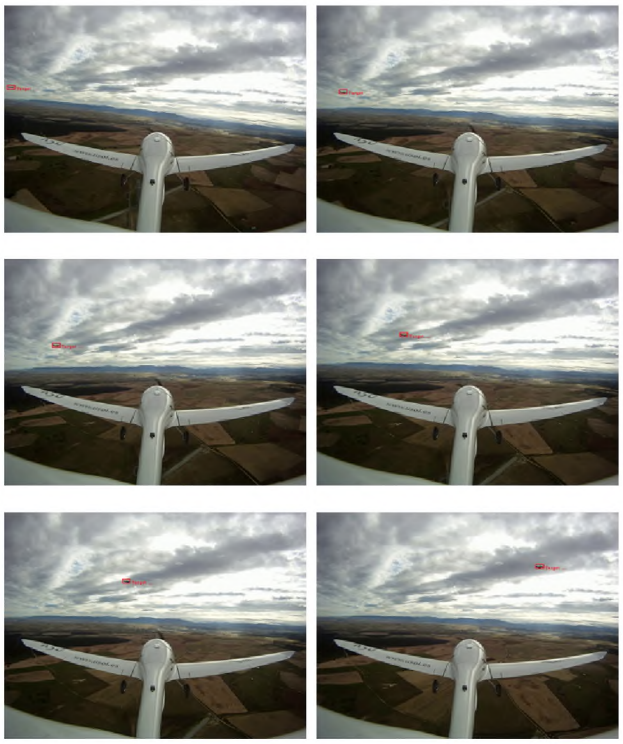

Fig. 12: Evaluation of AM3 tracking performance with a cloudy sky background (i.e. 1RLB737.avi). The tracker (red) can locate the aircraft during all the evaluation process under challenging situations.

ACKNOWLEDGMENT

The work reported in this paper is the consecution of several research stages at the Computer Vision Group-Universidad Politécnica de Madrid. This work has been sponsored by the Spanish Science and Technology Ministry under the grant CICYT DPI2010-20751-C02-01, the E-Vision Project (TSI-020100-2011-363) and the China Scholarship Council (CSC). The authors would like to thank the company USol as E-Vision's project coordinator for the valuable data and feedback provided.